Enhancing Microbial Genome Finishing With New Sequencing Technologies.

B. Foster1, K. Labutti1, E. Goltsman1, S. Trong1, S. Lowry1, A. Clum1, H. Sun1 and A. L. Lapidus1

1 Joint Genome Institute Production Genomics Facility, Walnut Creek, CA

Finishing is the process of enhancing a genome draft assembly into its final high quality state. It entails closing all gaps, resolving all repeats and mis-assemblies, and bringing final sequence quality to a pre-defined standard (polishing). Traditionally, this was achieved using Sanger clone based targeted sequencing reactions, sequencing targeted PCR products, as well as a host of additional techniques. This method, although effective, does have its drawbacks. Particular regions of the genome may be difficult to clone and can be under-represented in the constructed libraries and therefore also in the genome draft. Using only clone based sequencing technology to finish microbial genomes is time consuming, expensive and often a manual iterative process.

New technologies such as 454 and Solexa can resolve some of the limitations presented by the traditional clone based Sanger sequencing method. Although 454 pyrosequencing is currently limited in its ability to handle repetitive sequences, it does provide greater sequence coverage and is not subject to the same cloning bias as Sanger sequencing. Pyrosequencing is also helpful solving problems related to high GC secondary structures. Solexa, in addition, provides a very high volume of read sequence at fractions of the cost of traditional sequencing. Although the read length is short for Solexa (currently about 40 bp), the high sequence depth produced by this platform makes it ideal for assembled genome polishing. Boosting overall confidence in the genome sequence by Sanger sequencing alone can take as much as 50 percent of time and resources allocated to finish a microbial genome. Solexa based polishing reduces this time and cost significantly.

454 and Solexa used in combination with traditional clone based Sanger sequencing have greatly accelerated the microbial genome finishing process at JGI. Our current strategy of using 454 to assist in overall layout, Solexa coverage to enhance confidence in the genome basecalls, and Sanger sequencing to resolve repeats and other difficult genome specific problems, has lead to a greater capability to produce high quality finished sequence at a higher throughput and reduced cost.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

LBNL-63581 Abs.

UCRL-ABS-236414